Adv. Appl. Prob. **47,** 378–401 (2015)
Printed in Northern Ireland
© *Applied Probability Trust* 2015

AMERICAN OPTION VALUATION UNDER CONTINUOUS-TIME MARKOV CHAINS

B. ERIKSSON * ** AND
M. R. PISTORIUS,* *** *Imperial College London*

Abstract

This paper is concerned with the solution of the optimal stopping problem associated to the value of American options driven by continuous-time Markov chains. The value-function of an American option in this setting is characterised as the unique solution (in a distributional sense) of a system of variational inequalities. Furthermore, with continuous and smooth fit principles not applicable in this discrete state-space setting, a novel explicit characterisation is provided of the optimal stopping boundary in terms of the generator of the underlying Markov chain. Subsequently, an algorithm is presented for the valuation of American options under Markov chain models. By application to a suitably chosen sequence of Markov chains, the algorithm provides an approximate valuation of an American option under a class of Markov models that includes diffusion models, exponential Lévy models, and stochastic differential equations driven by Lévy processes. Numerical experiments for a range of different models suggest that the approximation algorithm is flexible and accurate. A proof of convergence is also provided.

Keywords: Markov chain; American option; free-boundary problem; optimal stopping; Feller process; numerical approximation

2010 Mathematics Subject Classification: Primary 91G20
Secondary 60J27; 65C40

1. Introduction

American options. The valuation of American options is an active research topic that has received a good deal of attention in the literature. Related American-type optimal stopping problems turn up in the modelling of trading and investment decisions, and real options (see, e.g. [4] and [6]). The theoretical and numerical aspects of American option valuation have been investigated using a diverse collection of tools, methods, and techniques, in several different settings; see [11] for an overview and references. It was understood early on that, as a consequence of the embedded optionality of the time of exercise, the value of an American option is equal to the value of an optimal stopping problem. For instance, under Samuelson's geometric Brownian motion model, which is considered to be the benchmark model for the evolution of the price of a risky stock, the optimal policy in the case of an American put is to exercise at the first moment the stock price falls below a certain boundary. In this setting it was first observed by McKean [22] that the value-function of an American option solves a free-boundary problem. Jacka [13] and Peškir [24] established this exercise boundary to be the unique solution

Received 12 June 2013; revision received 1 May 2014.
* Postal address: Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
** Email address: eriksson_bjorn@hotmail.com
*** Email address: m.pistorius@imperial.ac.uk

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

of an integral equation. Motivated by the observed features of empirical returns data the focus in modelling has subsequently shifted to more general classes of Markov processes, such as diffusions and jump processes. In the setting of Lévy processes the analytical characterisation of the value-function and optimal boundary of an American put was investigated among others by Boyarchenko and Levendorskiĭ [5] and Lamberton and Mikou [21]. In another line of research, going back at least as far as Cox *et al.* [10], a discrete-time and discrete-space approach was developed for the valuation of American options in the setting of a binomial tree. In later years many extensions and refinements of the discrete-time approach have been developed, for example, to tri- and multinomial trees. The connection between the two approaches was investigated in, e.g. Lamberton [19], Ahn and Song [2], and Szimayer and Maller [26] where (rates of) convergence of the values of American options under binomial and trinomial, and finite state models were established to those under the limiting Brownian or Lévy model, respectively. Kushner and Dupuis [18] proposed numerical methods for the solution of stochastic control problems in diffusion settings based on an approximation of the state process by Markov chains.

American options under Markov chains. In this paper we consider the optimal stopping problem associated to an American option in the setting of a continuous-time Markov chain with discrete state-space. Stochastic processes from this class have served as models for the evolution of random quantities that take values in lattices. Models from this class, which contains the classical birth–death processes, have recently also been deployed to model the state of the order book or the limit price; see, e.g. [1] and the references therein. Furthermore, Markov chains have been deployed as models on a discrete state-space that closely approximate continuous-space diffusions, jump-diffusions, and general Feller processes. In a continuous-time Markov chain setting, we solve the optimal stopping problem associated to the valuation of an American option with a payoff that is a function of the Markov chain. While it follows from the general theory of optimal stopping that the optimal stopping time is given by the first passage time into a certain set (see [25]), the characterisation of the value-function as the solution of a corresponding free-boundary problem and the identification of the optimal boundary involve nonstandard arguments. Taking advantage of the explicit form of the semigroup we demonstrate that the value-function of such an American option is the unique solution in a distributional sense of an associated free-boundary problem, and deduce that the value-function is in fact a classical solution by showing that it is continuously differentiable as function of time (see Theorem 3.1 below). In cases when the payoff and Markov process are sufficiently regular, pasting principles have been used to identify and characterise the optimal boundary. In the case that the payoff function is continuously differentiable in a neighbourhood of the boundary and the underlying is a real-valued Feller process, the general theory of optimal stopping (see [25]) suggests that it can be expected that, at the boundary, the value-function is continuously differentiable if, for the Feller process, the boundary is regular for itself, while the value-function can be expected to be merely continuous at the boundary if, for the Feller process, the boundary is irregular for itself. These two heuristics are known as the smooth-pasting and continuous-pasting principles, respectively. See [25] for a general treatment of pasting principles, and [3] and [20] for an investigation of the validity of pasting principles in the case of the optimal stopping problem associated to an American put option under a Lévy process. However, in the case of a discrete state-space with finite transition rates the smooth- and continuous-pasting principles no longer apply due to the lack of smoothness that is a result of the discrete state-space. In the absence of pasting principles, we derive an explicit characterisation of the optimal stopping boundary directly in terms of the infinitesimal generator of the Markov chain, in the case that the optimal stopping boundary is monotone (see Theorem 4.1 below).

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

Algorithm. Deploying this characterisation, we design an algorithm for the computation of the value-function of an American option under a continuous-time Markov chain model. By constructing the Markov chain such that it closely follows the evolution of a given Feller process (e.g. by using the construction from [23]), this algorithm, with the constructed Markov chain as input, provides a method for the valuation of American options under the Feller process in question. An advantage of the Markov chain model is its computational tractability: we demonstrate in this paper that the described algorithm provides an efficient and accurate method for the valuation of American options, and the computation of the optimal boundary, using the powerful tools of matrix-based computations. The idea of valuation using Markov chain approximation goes back at least as far as Kushner [17] in the case of diffusions, and was further developed in, e.g. [23]. To illustrate its effectiveness, we implemented the algorithm for a local-volatility model with jumps, and report results (such as estimates of the errors) in Section 7. We also give a proof of convergence of the approximation method.

Contents. The remainder of the paper is organised as follows. Section 2 contains preliminaries and notation that is used throughout the paper. Section 3 is devoted to the free-boundary problem associated to the American option driven by a continuous-time Markov chain and contains a characterisation of the optimal boundary, and in Section 5 an algorithm is presented for solving this free-boundary problem. Convergence of the algorithm is established in Section 6, and a number of numerical examples are analysed in Section 7. Appendix A and Appendix B contain the dynamic programming algorithm for valuing American options using Markov chains and the proof of Lemma 2.1 below.

2. Preliminaries

2.1. Setting: Markov chains

We next set the notation that will be used throughout the paper. Let X be a continuous-time, time-homogeneous Markov chain with discrete state-space $\mathbb{G} = \{x_i, i \in \mathbb{N}\}$ and generator matrix Λ, defined on some filtered probability space $(\Omega, \mathcal{G}, \boldsymbol{G}, \mathbb{P})$ where $\boldsymbol{G} = \{\mathcal{G}_t\}_{t\in[0,T]}$ denotes the completed right-continuous filtration generated by X. Assume that X is a Feller process with càdlàg paths (see [8, Section 2.2] for background), and denote the infinitesimal generator of X by Λ. To avoid explosion of the chain X in finite time, we assume that Λ has uniformly bounded elements according to the following assumption.

Assumption 2.1. *The infinitesimal generator Λ of X satisfies the condition*

$$\sup_{x\in\mathbb{G}} |\Lambda(x, x)| < \infty.$$

Denoting by $l^\infty(\mathbb{G})$ the collection of bounded real-valued functions with domain $\mathbb{G}$, we recall that the semigroup of X is equal to the collection $(P_t, t \in \mathbb{R}_+)$ of maps $P_t : l^\infty(\mathbb{G}) \to l^\infty(\mathbb{G})$ that is expressed in terms of the infinitesimal generator $\Lambda : l^\infty(\mathbb{G}) \to l^\infty(\mathbb{G})$ of X by

$$(P_t f)(x) = \sum_{y\in\mathbb{G}} P_t(x, y) f(y), \qquad t \in \mathbb{R}_+,\ x \in \mathbb{G},\ f \in l^\infty(\mathbb{G}),$$

$$P_t(x, y) = \mathbb{P}\{X_t = y \mid X_0 = x\} =: \mathbb{P}_x\{X_t = y\}, \qquad x, y \in \mathbb{G},$$

with $P_t = \exp(t\Lambda) = \sum_{n=0}^{\infty}(t^n/n!)\Lambda^n$ and $\Lambda^n = \Lambda^{n-1} \circ \Lambda$, i.e. $\Lambda^n f = \Lambda^{n-1}(\Lambda f)$ for any $f \in l^\infty(\mathbb{G})$. The infinitesimal generator Λ is given by

$$\Lambda f(x) = \sum_{y\in\mathbb{G}} \Lambda(x, y) f(y), \qquad \Lambda(x, y) = (\Lambda\delta_y)(x), \quad x \in \mathbb{G},\ f \in l^\infty(\mathbb{G}),$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

with $(1-\delta_y(x))\Lambda(x,y) \geq 0$ and $\sum_{z\in\mathbb{G}} \Lambda(x,z) = 0$, $x, y \in \mathbb{G}$, where δ_y is the Kronecker delta, which is the map on $\mathbb{G}$ that is equal to 1 if x and y are equal and 0 otherwise. In particular, it follows that the expected value of the payoff $\phi(X_T)$ at time T, where ϕ is an arbitrary map from the set $l^\infty(\mathbb{G})$, is given by

$$\mathbb{E}_{t,x}\{\phi(X_T)\} = \mathbb{E}_{0,x}\{\phi(X_{T-t})\} = (\exp((T-t)\Lambda)\phi)(x), \qquad x \in \mathbb{G},\ t \in [0,T], \tag{2.1}$$

where $\mathbb{E}_{t,x}\{\cdot\} = \mathbb{E}\{\cdot \mid X_t = x\}$ denotes the conditional expectation under the measure $\mathbb{P}$ conditioned on $\{X_t = x\}$. For a bounded function $f\colon [0,T]\times\mathbb{G} \to \mathbb{R}$ we also use the notation

$$(P_u f)(t,x) = (P_u f_t)(x), \qquad t, u \in [0,T],$$

where f_t is the map $f_t\colon \mathbb{G} \to \mathbb{R}$ given by $f_t(x) = f(t,x)$. Discounting at rate $r \geq 0$ can be incorporated by replacing the infinitesimal generator Λ by the subgenerator $\Lambda^{(r)}$ given by

$$\Lambda^{(r)} = \Lambda - r\mathbb{I},$$

where $\mathbb{I}\colon l^\infty(\mathbb{G}) \to l^\infty(\mathbb{G})$ is the identity map, so (2.1) generalises to

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-rT}\phi(X_T)\} = (\exp((T-t)\Lambda^{(r)})\phi)(x), \qquad x \in \mathbb{G},\ t \in [0,T]. \tag{2.2}$$

Remark 2.1. The Markov property of the chain X together with the identity in (2.2) imply that the discounted process $\{\mathrm{e}^{-rt}X_t,\ t \in \mathbb{R}_+\}$ is a martingale precisely, if we have

$$\mathbb{E}_{0,x}\{\mathrm{e}^{-rt}X_t\} = x \quad \text{for all } x \in \mathbb{G}.$$

2.2. Dynkin's lemma

Throughout this paper the following version of Dynkin's lemma will be frequently deployed in the analysis.

Lemma 2.1. *Assume that the function $F\colon [0,T]\times\mathbb{G} \to \mathbb{R}$ is bounded and that for any $x \in \mathbb{G}$, the map $t \mapsto F(t,x)$ is continuous with density $f(t,x)$ that is nonnegative for almost every $t \in [0,T]$. Then we have for any $t \in [0,T]$ and any $\boldsymbol{G}$-stopping time τ taking values in the interval $[t,T]$,*

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}F(\tau,X_\tau)\} = F(t,x) + \mathbb{E}_{t,x}\left\{\int_t^\tau \mathrm{e}^{-r(s-t)}(\overline{\Lambda}F)(s,X_s)\,\mathrm{d}s\right\} \tag{2.3}$$

with the map $\overline{\Lambda}F\colon [0,T]\times\mathbb{G} \to \mathbb{R}$ defined by

$$(\overline{\Lambda}F)(t,x) = f(t,x) + (\Lambda^{(r)}F)(t,x), \qquad t \in [0,T],\ x \in \mathbb{G}. \tag{2.4}$$

A proof is provided in Appendix B.

3. A Markov chain free-boundary problem

An American option with a payoff function given by ϕ and maturity $T > 0$, on an underlying with price process denoted by $X = \{X_t,\ t \in [0,T]\}$, is a derivative security that entitles its holder to receive the payoff $\phi(X_t)$ at any time t prior to the maturity T that he/she wishes to exercise the contract. The most common type of American options are the American call option with strike K, which has payoff $\phi(s) = (s-K)^+$ (with $x^+ = \max\{x, 0\}$ for $x \in \mathbb{R}$), and the

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

American put option with strike K, which has payoff given by $\phi(s) = (K - s)^+$. We assume that the payoff function $\phi\colon \mathbb{G} \to \mathbb{R}_+$ is nonnegative and satisfies the integrability condition

$$\mathbb{E}_{0,x}\Big\{\sup_{t\in[0,T]} \phi(X_t)\Big\} < \infty, \qquad x \in \mathbb{G}. \tag{3.1}$$

The value V_t^* of the American option at time $t \in [0, T]$ with payoff function ϕ is given by

$$V_t^* = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\{\mathrm{e}^{-r\tau}\phi(X_\tau) \mid \mathcal{G}_t\},$$

where $\mathcal{T}_{t,T}$ denotes the set of $\boldsymbol{G}$-stopping times taking values between t and T. The process $V^* = \{V_t^*, t \in [0, T]\}$ is called the *Snell-envelope* of the collection of discounted payoffs $\Pi = \{\mathrm{e}^{-rt}\phi(X_t),\ t \in [0, T]\}$: it is the smallest $\boldsymbol{G}$-supermartingale that is bounded below by Π. The Markov property of X implies that $V_t^* = V(t, X_t)$, where the value-function of the American option $V = \{V(t, x),\ t \in [0, T],\ x \in \mathbb{G}\}$ is given by

$$\begin{aligned} V(t,x) &= \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}\phi(X_\tau)\} && (3.2)\\ &= \sup_{\tau\in\mathcal{T}_{0,T-t}} \mathbb{E}_{0,x}\{\mathrm{e}^{-r\tau}\phi(X_\tau)\}, \qquad (t,x)\in[0,T]\times\mathbb{G}, && (3.3)\end{aligned}$$

where the second line is a consequence of the homogeneity of the Markov process X. According to the general theory of optimal stopping (see [25]), we have that the solution of the optimal stopping problem in (3.2) is expressed in terms of a stopping region $\mathbb{S}$ and a continuation region $\mathbb{C}$ given by

$$\mathbb{S} = \{(s, x) \in [0, T] \times \mathbb{G}\colon V(s, x) = \phi(x)\},$$
$$\mathbb{C} = \{(s, x) \in [0, T] \times \mathbb{G}\colon V(s, x) > \phi(x)\}.$$

In particular, $\tau_{\mathbb{S}}(t)$ given by

$$\tau_{\mathbb{S}}(t) = \inf\{s \in [t, T]\colon X_s \in \mathbb{S}\}$$

is a $\boldsymbol{G}$-stopping time in the set $\mathcal{T}_{t,T}$ that achieves the supremum in (3.2). By combining it with the strong Markov property of X, it follows that

$$\{\mathrm{e}^{-r(t\wedge\tau)}V(t\wedge\tau, X_{t\wedge\tau}),\ t\in[0,T]\} \tag{3.4}$$

is a martingale for $\tau = \tau_{\mathbb{S}}(0)$. We can decompose $\mathbb{S}$ as follows:

$$\mathbb{S} = \bigcup_{x\in\mathbb{G}} \mathbb{S}(x) \times \{x\}, \qquad \mathbb{S}(x) = \{s \in [t, T]\colon V(s, x) = \phi(x)\}.$$

In the following result two properties of the value-function and its generator are recorded that will be used later.

Proposition 3.1. *The following hold for the value-function V.*

(i) *For each $x \in \mathbb{G}$, the map $t \mapsto V(t, x)$ is decreasing and continuous.*

(ii) *For each $x \in \mathbb{G}$, the map $\Lambda^{(r)}\colon [0, T] \to \mathbb{R}$ given by $t \mapsto [\Lambda^{(r)} f_t](x)$ with $f_t(x) = V(t, x)$ is continuous and is decreasing when restricted to $\mathbb{S}(x)$.*

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

Proof of (i). Since, for any $s, t \in [0, T]$ with $t < s$, we have $\mathcal{T}_{0,T-t} \supseteq \mathcal{T}_{0,T-s}$ it follows from the representation in (3.3) that we have $V(t, x) \geq V(s, x)$ for each $x \in \mathbb{G}$. Lebesgue's dominated convergence theorem, the fact that ϕ satisfies the integrability condition in (3.1) and the triangle inequality imply that $V(t, x)$ is continuous as a function of t for any fixed $x \in \mathbb{G}$.

Proof of (ii). Since $\Lambda^{(r)}$ is a subgenerator, we have

$$\Lambda^{(r)}(h, g) \geq 0, \quad g \neq h, \qquad \Lambda^{(r)}(g, g) \leq 0, \quad g, h \in \mathbb{G},$$

so it follows that for any function f satisfying:

$$\text{for all } x \in \mathbb{G}\colon\ f(x) \geq 0, \quad \text{there exists } h \in \mathbb{G}\colon\ f(h) = 0, \tag{3.5}$$

we have that $(\Lambda^{(r)} f)(h)$ is nonnegative.

For any $t_1, t_2 \in [0, T]$, $t_2 \geq t_1$, and $g \in \mathbb{G}$ such that (t_1, g) and (t_2, g) are elements of $\mathbb{S}$, the function $f\colon \mathbb{G} \to \mathbb{R}$ given by $f(x) = V(t_1, x) - V(t_2, x)$ satisfies the conditions in (3.5), by virtue of the facts that $t \mapsto V(t, x)$ is decreasing (by (i)) and that we have $V(t_1, g) = V(t_2, g) = \phi(g)$ (by the definition of $\mathbb{S}$). Hence, we deduce that $\Lambda^{(r)}(V(t_1, g) - V(t_2, g))$ is nonnegative, which shows the stated monotonicity.

Since, for each $h \in \mathbb{G}$, we have $\Lambda^{(r)} V(t, h) = \sum_{g \in \mathbb{G}} \Lambda^{(r)}(h, g) V(t, g)$, it follows from the continuity of $t \mapsto V(t, g)$ (shown in (i)), the boundedness of V (by (3.1)), Assumption 2.1, and Lebesgue's dominated convergence theorem that $t \mapsto \Lambda^{(r)} V(t, h)$ is also continuous.

The monotonicity of $t \mapsto V(t, x)$ stated in Proposition 3.1(i) implies that if a point (t, x) lies in $\mathbb{S}$ then any point of the form (s, x) for $s > t$ also lies in $\mathbb{S}$. Thus, since $t \mapsto V(t, x)$ is continuous, the set $\mathbb{S}(x)$ is closed and is of the form

$$\mathbb{S}(x) = [\tau(x), T] \quad \text{for some } \tau(x) \in [0, T].$$

Associated to the value-function of the American option is the system of variational inequalities given by

$$\Lambda_t V(t, x) \leq 0 \quad \text{for } (t, x) \in [0, T] \times \mathbb{G}, \tag{3.6}$$

$$\Lambda_t V(t, x) = 0 \quad \text{for } (t, x) \in \mathbb{C}, \tag{3.7}$$

$$V(t, x) = \phi(x) \quad \text{for } (t, x) \in \mathbb{S}, \tag{3.8}$$

$$V(t, x) > \phi(x) \quad \text{for } (t, x) \in \mathbb{C}, \tag{3.9}$$

where Λ_t denotes the infinitesimal generator of the time-space process (t, X_t), which acts on functions F in the set $C^1([0, T] \times \mathbb{G})$ (the set of functions $F\colon [0, T] \times \mathbb{G} \to \mathbb{R}$ that are continuously differentiable as a function of the first argument), as follows:

$$\Lambda_t F = \frac{\partial F}{\partial t} + \Lambda^{(r)} F.$$

Since *a priori* we know only that the value-function V is continuous and decreasing as a function of t, V may not be a classical solution of the system in (3.6)–(3.9) of variational inequalities. A function $V\colon [0, T] \times \mathbb{G} \to \mathbb{R}$ is called a solution in a *distributional sense* of the system in (3.6)–(3.9) if V satisfies (3.6)–(3.9) with the map $\Lambda_t V$ replaced by the map $\overline{\Lambda} V$ that was defined in (2.4).

We have the following theorem for the existence and uniqueness.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

Theorem 3.1. *The function V defined in (3.2) is the unique continuous decreasing function that solves the system of variational inequalities in (3.6)–(3.9) in a distributional sense.*

Furthermore, we have

$$(\Lambda^{(r)}V)(\tau(x), x) = 0 \quad \textit{for any } x \in \mathbb{G} \textit{ satisfying } \tau(x) < T, \tag{3.10}$$

$$(\Lambda^{(r)}V)(t, x) \le 0 \quad \textit{for any } x \in \mathbb{G} \textit{ and } t \in [0, T] \textit{ with } t > \tau(x). \tag{3.11}$$

In particular, the value-function V is a classical solution of the system in (3.6)–(3.9).

Proof of Theorem 3.1. (i) Existence. That V is decreasing and continuous follows from Proposition 3.1. We show that V satisfies (3.8) and (3.9), and satisfies (3.6) and (3.7) in a distributional sense. Note that (3.8) and (3.9) hold true by the definition of the stopping and continuation regions $\mathbb{S}$ and $\mathbb{C}$. Next, we verify that (3.6) holds. Since $t \mapsto V(t, x)$ is decreasing and continuous, $V(\cdot, x)$ admits a density that is almost everywhere nonpositive. For any $x \in \mathbb{G}$ and any $t \in [0, T]$ and any stopping time $\tau \in \mathcal{T}_{t,T}$ we have, by Lemma 2.1 (Dynkin's lemma),

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}V(\tau, X_\tau)\} = V(t, x) + \mathbb{E}_{t,x}\bigg\{\int_t^\tau \mathrm{e}^{-r(s-t)}(\overline{\Lambda}_t V)(s, X_s)\,\mathrm{d}s\bigg\}, \tag{3.12}$$

where $\overline{\Lambda}_t V$ is defined in (2.4). As the discounted value-process $\mathrm{e}^{-rt}V(t, X_t)$ is a supermartingale, we have for any pair $t_1, t_2 \in [0, T]$ with $t_1 < t_2$ and any $x \in \mathbb{G}$ the inequality $\mathbb{E}_{t_1,x}\{\mathrm{e}^{-rt_2}V(t_2, X_{t_2})\} \le \mathrm{e}^{-rt_1}V(t_1, x)$ which yields in view of (3.12) the relation

$$B(t_1, t_2, x_1) := \mathbb{E}_{t_1,x}\bigg\{\int_{t_1}^{t_2} \mathrm{e}^{-r(s-t_1)}\overline{\Lambda}_t V(s, X_s)\,\mathrm{d}s\bigg\} \le 0. \tag{3.13}$$

To see that (3.13) implies that (3.6) is satisfied (in a distributional sense), note that the left-hand side of (3.13) is equal to

$$B(t_1, t_2, x_1) = \sum_{y \in \mathbb{G}} \int_{t_1}^{t_2} \mathrm{e}^{-r(s-t_1)}\overline{\Lambda}_t V(s, y)\mathbb{P}_{x,t_1}\{X_s = y\}\,\mathrm{d}s.$$

Since we have $\mathbb{P}_{x,t_1}\{X_s \ne x\} = -\Lambda(x, x)(s - t_1) + o(t_2 - t_1)$ $(t_2 \searrow t_1)$ for all $s \le t_2$ (as X is a continuous-time Markov chain), it follows that $\overline{\Lambda}_t V(s, y)$ is nonpositive for almost every $t \in [0, T]$ and for all $y \in \mathbb{G}$. Thus, the claim follows from (3.13).

Finally, we check that (3.7) is satisfied. Since the stopped process $\mathrm{e}^{-r(t\wedge\tau_{\mathbb{S}})}V(t \wedge \tau_{\mathbb{S}}, X_{t\wedge\tau_{\mathbb{S}}})$ is a $\mathbb{P}_{t,x}$-martingale for any $(t, x) \in \mathbb{C}$ (cf. (3.4)), it follows that for any $t_1, t_2 \in [0, T]$ with $t_1 < t_2$,

$$\mathbb{E}_{t_1,x}\{\mathrm{e}^{-r(t_2\wedge\tau_{\mathbb{S}})}V(t_2 \wedge \tau_{\mathbb{S}}, X_{t_2\wedge\tau_{\mathbb{S}}})\} = \mathbb{E}_{t_1,x}\{\mathrm{e}^{-r(t_1\wedge\tau_S)}V(t_1 \wedge \tau_{\mathbb{S}}, X_{t_1\wedge\tau_{\mathbb{S}}})\},$$

which is equal to $\mathrm{e}^{-rt_1}V(t_1, x)$ so that, in view of the equality in (3.12), we have the equality

$$\mathbb{E}_{t_1,x}\bigg\{\int_{t_1}^{t_2\wedge\tau_{\mathbb{S}}} \mathrm{e}^{-r(s-t_1)}\overline{\Lambda}_t V(s, X_s)\,\mathrm{d}s\bigg\} = 0.$$

A line of reasoning that is similar to the one used in the previous paragraph shows that $\overline{\Lambda}_t V(t, x) = 0$ for almost every $t \in [0, T]$ and every $x \in \mathbb{G}$ with $(t, x) \in \mathbb{C}$, so we deduce that (3.7) holds (in a distributional sense).

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

(ii) Uniqueness. Assume that $\widetilde{V}$ is a continuous decreasing function that solves the system in (3.6)–(3.9) in a distributional sense. An application of Lemma 2.1 shows that for any stopping time $\tau \in \mathbb{T}_{t,T}$, we have

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}\phi(X_\tau)\} \le \mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}\widetilde{V}(\tau, X_\tau)\} \le \widetilde{V}(t,x), \tag{3.14}$$

where we use (3.6), (3.8), and (3.9). Taking the supremum in (3.14) over $\tau \in \mathcal{T}_{t,T}$ shows that $V(t,x) \le \widetilde{V}(t,x)$. Similarly, an application of Dynkin's lemma shows that if the function $\widetilde{V}$ solves the system in (3.7) and (3.8) in a distributional sense then we have

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau_{\mathbb{S}}-t)}\phi(X_{\tau_{\mathbb{S}}})\} = \widetilde{V}(t,x), \qquad (t,x) \in [0,T]\times\mathbb{G}.$$

Hence, choosing $\tau = \tau_{\mathbb{S}}$ in (3.14) turns the inequalities into equalities and it follows that $V(t,x) = \widetilde{V}(t,x)$. We deduce that the solution of the system in (3.6)–(3.9) is unique in a distributional sense.

(iii) Equations (3.10) and (3.11). Since $t \mapsto V(t,x)$ is decreasing (Proposition 3.1), we have that $V(\cdot,x)$ admits a density that is nonpositive for almost every $t \in [0,T]$ and any $x \in \mathbb{G}$ with $(t,x) \in \mathbb{C}$. Hence, in combination with the equality in (3.7) and the continuity of $t \mapsto \Lambda^{(r)}V(t,x)$, we have

$$\Lambda^{(r)}V(t,x) \ge 0 \quad \text{for all } (t,x) \in \mathbb{C}.$$

Observing that the map $t \mapsto V(t,x)$ restricted to the interval $\mathbb{S}(x) = [\tau(x), T]$ is a constant equal to $\phi(x)$, we see that the density of $V(\cdot,x)$ is equal to 0 for almost every $t \in [\tau(x),T]$ and $x \in \mathbb{G}$ for which $\tau(x)$ is strictly smaller than T. Thus, in view of the relation in (3.6) and the continuity of the map $t \mapsto \Lambda^{(r)}V(t,x)$, we have

$$0 \ge \Lambda^{(r)}V(t,x) \quad \text{for any } t \in [\tau(x), T]$$

and $x \in \mathbb{G}$ with $\tau(x) < T$. Since the map $t \mapsto \Lambda^{(r)}V(t,x)$ is continuous, nonnegative for $t < \tau(x)$ and nonpositive for $t > \tau(x)$, the intermediate value theorem implies that $\Lambda^{(r)}V(\tau(x),x)$ is equal to 0, and the proofs of (3.10) and (3.11) are complete. The proof of the fact that V is a classical solution is given in the next section.

4. Characterisation of the optimal boundary

In this section we present a characterisation of the stopping region $\mathbb{S}$. To simplify the presentation we will make the following assumption throughout this and the next section.

Assumption 4.1. *The stopping region is of the form*

$$\mathbb{S} = \{(t,x) \in [0,T]\times\mathbb{G} : x \le B(t)\},$$

where the optimal boundary $t \mapsto B(t)$ is increasing as a function of time t with $B(T)$ taking a finite value.

If the sequences $\mathcal{X} = \{x_1, x_2, \dots\}$ and $\{\tau(x_1), \tau(x_2), \dots\}$ are nondecreasing, then the optimal boundary is given by $B(t) = \sup\{x_i \in \mathcal{X} : t \in [\tau(x_i), T]\}$. This form of B is, for example, encountered in the case of an American put option under a continuous-time Markov chain model that is spatially homogeneous (see Figure 1).

Denote by

$$\mathbb{B} = \{B(\tau(x)),\ x \in \mathbb{G}\} = \{b_i\}_i, \qquad b_i > b_{i+1},$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

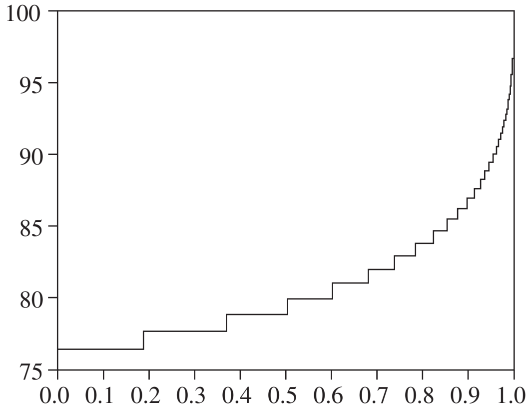

FIGURE 1: The optimal boundary corresponding to an at-the-money American put option with strike $S_0 = K = 100$ and maturity $T = 1$ when interest rate and dividend yield are given by $r = 0.1$ and $\delta = 0$ and the underlying price process is given by a Markov chain that closely approximates a geometric Brownian motion with volatility $\sigma = 0.3$. The chain has a state-space of size 200 and is constructed by matching the instantaneous moments of the Markov chain with those of the Brownian motion, using the procedure described in [23].

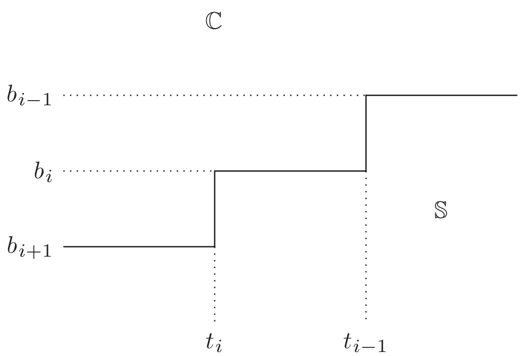

FIGURE 2: A close-up of the optimal boundary, illustrating the values b_i and t_i.

the set of distinct elements in $\{B(\tau(x)),\ x \in \mathbb{G}\}$ that the optimal boundary takes (in order of decreasing magnitude or, equivalently, increasing time to maturity T; see Figure 2) and by

$$t_i = \tau(b_i), \qquad i = 1, 2, \dots,$$

the first epoch t in the interval $[0, T]$ that the optimal boundary $B(t)$ is equal to b_i. At this point we note that (a) the sequence $\{t_i\}_i$ is decreasing and (b) the boundary B is constant between the epochs of t_i and has a discontinuity at the epochs of t_i. Given the times t_i and the optimal barrier levels b_i, the American option can be valued recursively: the value-function V of the American option is equal to the value-function of a barrier option contract with time-dependent barrier B that entitles the holder to a rebate payment $\phi(X_{\tau_B})$ if the epoch $\tau_B = \inf\{t \geq 0\colon X_t \leq B(t)\}$ is strictly smaller than T and to a payment $\phi(X_T)$ in the case that the epoch τ_B is larger or equal to T. From the Markov property of X applied at the epochs t_i and that the barrier B has jumps,

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

it follows that the function V is equal to the final value V_{N^*} of the following recursion:

$$\begin{aligned} V_i(t,x) &= \mathbb{E}_{t,x}\{\mathrm{e}^{-rT_{b_i}\circ\theta_t}\phi(X_{t+T_{b_i}\circ\theta_t})\ \mathbf{1}_{\{T_{b_i}\circ\theta_t<t_{i-1}-t\}} \\ &\qquad + \mathrm{e}^{-r(t_{i-1}-t)}V_{i-1}(t_{i-1},X_{t_{i-1}})\,\mathbf{1}_{\{T_{b_i}\circ\theta_t>t_{i-1}-t\}}\} \\ &= \mathbb{E}_{t,x}\{\mathrm{e}^{-r(T_{b_i}\circ\theta_t\wedge(t_{i-1}-t))}V_{i-1}(t_{i-1},X_{(t+T_{b_i}\circ\theta_t)\wedge t_{i-1}})\} \end{aligned} \tag{4.1}$$

for $t\in[0,t_{i-1}]$ and all $i\geq 1$ and $x\in\mathbb{G}$, with $V_0(t,x)=\phi(x)$ for $t\in[0,T]$, where $T_{b_i}=\inf\{s\geq 0\colon X_s\leq b_i\}$ and θ_t denotes the shift-operator (defined by $\theta_t(\omega)=\omega(t+\cdot)$ for all $\omega\in\Omega$), so $T_{b_i}\circ\theta_t=\inf\{s\geq 0\colon X_{t+s}\leq b_i\}$ holds. We denote by $\mathbf{1}$ the indicator function. Note that we have

$$V(t,x)=V_i(t,x)=\phi(x)\quad\text{for any pair }(x,t)$$

with $x\in\mathbb{S}$ and $t\leq t_{i-1}$,

$$V(t,x)=V_i(t,x)\quad\text{for any }t\in[t_i,t_{i-1}]$$

and $x\in\mathbb{G}$. Thus, the optimal value-function V is equal to V_i on the time interval $[t_i,t_{i-1}]$.

Next, we will characterise the collection of epochs $\{t_i\}_i$ in terms of the value of the time-space generator Λ_t applied to the functions V_i.

Theorem 4.1. *Let V_i be defined by (4.1). For any $i\in\mathbb{N}$ with $b_i\in\mathbb{B}$ and $t_i<T$, it holds*

$$\Lambda_t V_i(t,x)=0\quad\text{for } x>b_i,\ t\in[0,t_{i-1}), \tag{4.2}$$

$$\Lambda_t V_i(t,x)=\Lambda^{(r)}V_i(t,x)=0\quad\text{for } x=b_i,\ t=t_i, \tag{4.3}$$

$$\Lambda_t V_i(t,x)=\Lambda^{(r)}V_i(t,x)\leq 0\quad\text{for } x\leq b_i,\ t_i<t. \tag{4.4}$$

$$\Lambda_t V_i(t,x)=\Lambda^{(r)}V_i(t,x)>0\quad\text{for } x=b_i,\ t<t_i, \tag{4.5}$$

The proof is based on the following auxiliary lemma.

Lemma 4.1. *For any $i\in\mathbb{N}$ with $b_i\in\mathbb{B}$ and any $x\in\mathbb{G}$, the function $V_i(\cdot,x)\colon[0,t_{i-1}]\to\mathbb{R}$ given by $t\mapsto V_i(t,x)$ is decreasing and continuous. As a consequence, the function $\Lambda^{(r)}V_i(\cdot,b_i)\colon[0,t_{i-1}]\to\mathbb{R}$ given by $t\mapsto(\Lambda^{(r)}V_i)(t,b_i)$ is continuous and decreasing on $[0,t_{i-1}]$.*

Proof. Let $x\in\mathbb{G}$ and i with $b_i\in\mathbb{B}$ be arbitrary and given. The function $t\mapsto V_i(t,x)$ restricted to the interval (t_i,t_{i-1}) is equal to the function $t\mapsto V(t,x)$, which was shown to be decreasing from Proposition 3.1. We next turn to the $t\leq t_i$ case. Note that for any $t\in[0,t_{i-1}]$ we have that $V_i(t,x)$ is equal to

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(T_{b_i}\circ\theta_t\wedge(t_{i-1}-t))}V_{i-1}(t_{i-1},X_{(t+T_{b_i}\circ\theta_t)\wedge t_{i-1}})\}=(\exp[(t_{i-1}-t)\widetilde{\Lambda}_r^{(i)}]\phi_{i-1})(x), \tag{4.6}$$

where $\phi_{i-1}\colon\mathbb{G}\to\mathbb{R}$ is given by $\phi_{i-1}(x)=V_{i-1}(t_{i-1},x)$ and $\widetilde{\Lambda}_r^{(i)}$ is the subgenerator of X, discounted at rate r and stopped upon first entrance into the set $\{x\in\mathbb{G}\colon x\leq b_i\}$ (see (5.1) below).

Thus, for any $t,s\in[0,t_{i-1}]$ with $t>s$ we have

$$V_i(s,x)-V_i(t,x)=[\exp[(t_{i-1}-t)\widetilde{\Lambda}_r^{(i)}](\exp[(t-s)\widetilde{\Lambda}_r^{(i)}]-I)\phi_{i-1}](x), \tag{4.7}$$

where I denotes the identity. Since $t\mapsto V_i(t,x)$ is decreasing for $t\in[t_i,t_{i-1}]$, we deduce from (4.6) and (4.7),

$$V_i(t,x)-V_i(t_{i-1},x)=[(\exp[(t_{i-1}-t)\widetilde{\Lambda}_r^{(i)}]-I)\phi_{i-1}](x)\geq 0 \tag{4.8}$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

for any $t \in [t_i, t_{i-1}]$ and $x \in \mathbb{G}$. In view of (4.7) and (4.8), it follows that

$$V_i(t, x) - V_i(s, x) \le 0 \quad \text{for any } s, t \in [0, t_{i-1}] \tag{4.9}$$

with $t - s \in [0, t_{i-1} - t_i]$. As the difference $t_{i-1} - t_i$ is strictly positive, the statement in (4.9) implies that $V_i(t, x) - V_i(s, x) \le 0$ for any $s, t \in [0, t_{i-1}]$ with $t \ge s$. The proof of the monotonicity of $V_i(\cdot, x)$ is complete. The continuity of $t \mapsto V_i(t, x)$ for any $x \in \mathbb{G}$ follows from the continuity of the semigroup associated to the subgenerator $\widetilde{\Lambda}_r^{(i)}$, while the continuity of $t \mapsto (\Lambda^{(r)} V_i)(t, b_i)$ follows by an application of Lebesgue's dominated convergence theorem, which is justified in view of the continuity of $t \mapsto V_i(t, x)$, the boundedness of V_i, and Assumption 2.1.

By an argument that is analogous to the one deployed in the proof of Proposition 3.1(ii) (noting that $V_i(t, b_i) = \phi(b_i)$ for any $t \in [0, t_{i-1}]$), it follows that the monotonicity of $V_i(\cdot, x)$ implies the monotonicity of $t \mapsto (\Lambda^{(r)} V_i)(t, b_i)$ on the interval $[0, t_{i-1}]$.

Proof of Theorem 3.1 (continued). (iv) *Classical solution.* We start by noting that Assumption 4.1 does not play any other role in the proof than simplifying the notation and definitions (of, e.g. the functions V_i), and the proof in the general case is obtained by a straightforward adaptation of the proof that follows below. To show that V is a classical solution, it suffices to show that at every t in $[0, T]$ and x in $\mathbb{G}$ the map $t \mapsto V(t, x)$ is continuously differentiable. Noting that the restrictions of the functions V and V_i to the interval (t_i, t_{i-1}) are equal, we deduce that V is continuously differentiable at every t in (t_i, t_{i-1}) with derivative given by

$$\frac{\partial V}{\partial t}(t, x) = \frac{\partial V_i}{\partial t}(t, x) = -(\widetilde{\Lambda}_r^{(i)} V)(t, x), \qquad t \in (t_i, t_{i-1}),\ (t, x) \in \mathbb{C}.$$

Furthermore, since the function V is a solution of the system of variational equalities in (3.6)–(3.9) and is constant as a function of t in the stopping region $\mathbb{S}$ it follows that

$$\frac{\partial V}{\partial t}(t, x) = -(\Lambda^{(r)} V)(t, x) \quad \text{for any } t \in (t_i, t_{i-1}) \tag{4.10}$$

with $(t, x) \in \mathbb{C}$,

$$\frac{\partial V}{\partial t}(t, x) = 0 \quad \text{for any pair } (t, x) \tag{4.11}$$

with $t \in [t_i, t_{i-1}]$ and $(t, x) \in \mathbb{S}$. Here, we used that for any $x \in \mathbb{G}$, the definition of the sequence $(t_i)_i$ implies that if there exists a $t \in (t_i, t_{i-1})$ with $(t, x) \in \mathbb{S}$ then we have $(t, x) \in \mathbb{S}$ for all $t \in [t_i, t_{i-1}]$. To complete the proof of the continuous differentiability of V we finally consider the $t = t_i$ case. If t_i is such that (t_i, x) is an element of the continuation region $\mathbb{C}$ then it follows from the expression in (4.10) and the fact that the continuation region is open that the left limit and right limit of $(\partial V/\partial t)(t, x)$ at t_i are equal. If t_i is such that (t_i, x) is an element of the stopping region $\mathbb{S}$ then we have $\tau(x)$ is smaller or equal to t_i. In the case $\tau(x) < t_i$ it follows from (4.11) that the right- and left limit of $(\partial V/\partial t)(t, x)$ at t_i are equal to 0. In the $\tau(x) = t_i$ case, we note that the right limit is equal to 0, while the left limit of $(\partial V/\partial t)(t, x)$ at t_i is equal to $(\Lambda^{(r)} V)(\tau(x), x)$ which, in view of (3.10), is also equal to 0. Thus, we deduce that at all $t \in [0, T]$ and $x \in \mathbb{G}$ the function V is continuously differentiable and the proof is complete.

Proof of Theorem 4.1. Since $V_i(t, x) = V(t, x)$ for $t \in [t_i, t_{i-1}]$, (4.3) and (4.4) hold in view of Theorem 3.1.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

The function V_i is the value-function of a down-and-out barrier option with maturity t_{i-1} rebate $\phi(x)$ and terminal payoff function $V_{i-1}(t_{i-1}, x)$. Since the process $\mathrm{e}^{-r(t\wedge t_{i-1}\wedge T_{b_i})} V_i(t \wedge t_{i-1} \wedge T_{b_i}, X_{t\wedge t_{i-1}\wedge T_{b_i}})$ is a martingale, it follows by an analogous reasoning as the one used in the proof of Theorem 3.1 that we have $\Lambda_t V_i(t, x) = 0$ for $x > b_i$ and $t < t_{i-1}$. Hence, (4.2) holds.

Finally, we turn to the proof of (4.5). We start by observing that $(\Lambda^{(r)} V_i)(t, b_i)$ is nonnegative on the interval $t \in [0, t_i]$ in view of Lemma 4.1 and (4.3). We next show that $(\Lambda^{(r)} V_i)(t, b_i)$ is in fact strictly positive on the interval $[0, t_i)$.

By an application of Dynkin's lemma, Lemma 2.1, we obtain

$$\begin{aligned} V_i(t, x) - V_{i+1}(t, x) &= \mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau-t)}\{V_i(\tau, X_\tau) - V_{i+1}(\tau, X_\tau)\}\} \\ &\quad - \mathbb{E}_{t,x}\left\{\int_t^\tau \mathrm{e}^{-r(s-t)}\{\Lambda_t V_i(s, X_s) - \Lambda_t V_{i+1}(s, X_s)\}\,\mathrm{d}s\right\} \end{aligned} \tag{4.12}$$

for all $x \in \mathbb{G}$, $t \le t_i$ and $\tau \in \mathcal{T}_{t,t_i}$. Since by (4.2) we have

$$\Lambda_t V_i(s, x) = 0 \quad \text{for any } x > b_i,$$

$s \in [0, t_{i-1})$ and any $i \in \mathbb{N}$ with $b_i \in \mathbb{B}$, and the collection $\{b_i\}_i$ is decreasing, choosing τ in (4.12) to be equal to

$$\tau_i = \min\{t + T_{b_{i+1}} \circ \theta_t, t_i\}$$

shows that the right-most expectation in (4.12) is equal to

$$\begin{aligned} &\mathbb{E}_{t,x}\left\{\int_t^{\tau_i} \mathrm{e}^{-r(s-t)}\{\Lambda_t V_i(s, X_s) - \Lambda_t V_{i+1}(s, X_s)\}\,\mathrm{d}s\right\} \\ &\qquad = \mathbb{E}_{t,x}\left\{\int_t^{\tau_i} \mathrm{e}^{-r(s-t)}\Lambda_t V_i(s, X_s)\,\mathbf{1}_{\{X_s=b_i\}}\,\mathrm{d}s\right\}. \end{aligned} \tag{4.13}$$

Furthermore, we have that $V_i(\tau_i, X_{\tau_i}) = V_{i+1}(\tau_i, X_{\tau_i})$ for the following two reasons: (a) it holds that $V_{i+1}(t_i, X_{t_i}) = V_i(t_i, X_{t_i})$ by definition of V_{i+1} and (b) we have that, on the set $\{t + T_{b_{i+1}} \circ \theta_t < t_i\}$,

$$V_i(t + T_{b_{i+1}} \circ \theta_t, X_{t+T_{b_{i+1}}\circ\theta_t}) = V_{i+1}(t + T_{b_{i+1}} \circ \theta_t, X_{t+T_{b_{i+1}}\circ\theta_t}) = \phi(X_{t+T_{b_{i+1}}\circ\theta_t})$$

as it holds that $X_{T_{b_{i+1}}} \le b_{i+1} < b_i$ by the definition of $T_{b_{i+1}}$ and the fact that b_i is decreasing as a function of i. Hence, we deduce the identity

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau_i-t)} V_i(\tau_i, X_{\tau_i})\} = \mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau_i-t)} V_{i+1}(\tau_i, X_{\tau_i})\}. \tag{4.14}$$

Combining (4.12), (4.13), and (4.14) shows that

$$V_i(t, x) - V_{i+1}(t, x) = \mathbb{E}_{t,x}\left\{\int_t^{\tau_i} \mathrm{e}^{-r(s-t)}\Lambda_t V_i(s, X_s)\,\mathbf{1}_{\{X_s=b_i\}}\,\mathrm{d}s\right\}. \tag{4.15}$$

On the one hand, the construction of the value-functions $\{V_i\}$ and the definition of the collection $\{b_i\}$ imply that

$$V_{i+1}(t, b_i) > \phi(b_i) = V_i(t, b_i), \qquad t \in [0, t_i), \tag{4.16}$$

while, on the other hand, the equality $V_i(t, b_i) = \phi(b_i)$ for all $t \in [0, t_{i-1}]$ implies that $\partial V_i(t, b_i)/\partial t = 0$ for $t \in (0, t_{i-1})$, so we have

$$(\Lambda_t V_i)(t, b_i) = (\Lambda^{(r)} V_i)(t, b_i) \qquad t \in (0, t_{i-1}). \tag{4.17}$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

Thus, from (4.15), (4.16), and (4.17), we deduce that

$$\mathbb{E}_{t,b_i}\left\{\int_t^{\tau_i} \mathrm{e}^{-r(s-t)}\Lambda^{(r)}V_i(s,b_i)\,\mathbf{1}_{\{X_s=b_i\}}\,\mathrm{d}s\right\} > 0 \quad \text{for any } t \in [0, t_i]. \tag{4.18}$$

Since the map $t \mapsto \Lambda^{(r)}V_i(t, b_i)$ is continuous and nonnegative on the interval $[0, t_i]$ and it is straightforward to check that (4.18) remains valid with τ_i replaced by $\tau_i \wedge (t+u)$ for any $u > 0$, it follows that $\Lambda^{(r)}V_i(t, b_i) > 0$ for any $t \in [0, t_i)$.

5. Valuation algorithm

The characterisation of the free-boundary given in Theorem 4.1 can be deployed to compute the optimal boundary and the corresponding value of an American option under the Markov chain model. For the presentation of a valuation algorithm, we will restrict ourselves in this section to Markov chains with a finite state-space (of size N, say).

To identify the epochs $\{t_i\}$ a numerical method has to be deployed since the equations

$$\Lambda_t V_i(t, b_i) = 0,$$

are highly nonlinear in t. Except in degenerate cases, one may expect the map $s \mapsto \Lambda_t V_i(s, b_i)$ to be strictly decreasing, in which case the equation $(\Lambda_t V_i)(t, b_i) = 0$ admits a unique solution and it is efficient to use a solver such as the Newton–Raphson method (which is the method that was used in the examples in Section 7). (Note that although we could have attempted to compute t_i as a root of the function $s \mapsto \Lambda^{(r)}V_i(s, b_i)$, we found that working with $s \mapsto \Lambda_t V_i(s, b_i)$ yielded a more efficient numerical implementation). A procedure for computation of the value-function of an American option under a Markov chain model based on a solution of the corresponding free-boundary problem that was outlined in the previous paragraph is described in Algorithm 1 below. In order to be able to formulate the algorithm, we fix some extra notation. After relabelling, we may assume without loss of generality that the elements of the state-space $\mathbb{G} = \{x_i,\ i = 1, \ldots, N\}$, where N is the number of states, are ordered in decreasing order

$$x_N < x_{N-1} < \cdots < x_2 < x_1,$$

and we denote by

$$\mathbb{G}_{i:j} = \{x_k, k \in \{i, \ldots, j\}\} \qquad i < j, \quad i, j = 1, \ldots, N,$$

the slice of the state-space consisting of the elements $x_i, \ldots, x_j$. Furthermore, for any $i = 1, \ldots, N$, denote by $\widetilde{\Lambda}_r^{(i)}$ and $\overline{\Lambda}_r^{(i)}$ the (sub)generator matrices that can be obtained directly from the generator matrix $\Lambda^{(r)}$ as follows: (a) the pair satisfies

$$\widetilde{\Lambda}_r^{(i)} + \overline{\Lambda}_r^{(i)} = \Lambda^{(r)},$$

where we recall that $\Lambda^{(r)} = \Lambda - r\mathbb{I}$, and (b) $\widetilde{\Lambda}_r^{(i)}(x, y)$ is equal to $\Lambda^{(r)}(x, y)$ for $x, y \in \mathbb{G}_{1:i}$ and 0 for $x, y \in \mathbb{G}_{i+1:N}$, namely,

$$\widetilde{\Lambda}_r^{(i)}(x, y) = \begin{cases} \Lambda(x, y) - r & \text{for } x \in \mathbb{G}, x \geq x_i,\ x = y, \\ \Lambda(x, y) & \text{for } x, y \in \mathbb{G}, x \geq x_i,\ x \neq y, \\ 0 & \text{for } x \leq x_{i+1},\ x,\ y \in \mathbb{G}. \end{cases} \tag{5.1}$$

The matrix $\widetilde{\Lambda}_r^{(i)}$ is the generator matrix of a Markov chain that has the same law as the chain X that is stopped upon the first entrance into the set $\mathbb{G}_{i+1:N}$. The role of these matrices in barrier option valuation in Markov chain models is reviewed in Remark 5.1(ii) below.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

Algorithm 1. (*Markov chain free-boundary algorithm.*)
find index i of largest grid point $x_i \in \mathbb{G}$ such that $(\Lambda^{(r)}\phi)(x_i) < 0$
set $t^* \leftarrow T$
while $t^* > 0$
    find $s < t^*$ such that $\overline{\Lambda}_r^{(i)} \exp((t^* - s)\widetilde{\Lambda}_r^{(i)}\mathrm{i})\phi(x_i) = 0$;
    if $s > 0$
        set $\phi \leftarrow \exp((t^* - s)\widetilde{\Lambda}_r^{(i)})\phi$;
    else if $s \le 0$
        set $\phi \leftarrow \exp(t^*\widetilde{\Lambda}_r^{(i)})\phi$;
    set $i \leftarrow i + 1$; set $t^* \leftarrow s$;
end
return ϕ

Remark 5.1. In Algorithm 1 we used the following two facts.

(i) In view of the definition of the matrix $\overline{\Lambda}_r^{(i)}$ and the relation $(\mathrm{d}/\mathrm{d}t)\exp(tA) = A\exp(tA)$ that holds for any square matrix A, we have the equality

$$(\Lambda_t \exp((t^* - t)\widetilde{\Lambda}_r^{(i)}))|_{t=s} = O \iff \Lambda^{(r)} \exp((t^* - s)\widetilde{\Lambda}_r^{(i)}) = \widetilde{\Lambda}_r^{(i)} \exp((t^* - s)\widetilde{\Lambda}_r^{(i)}),$$

where O denotes a zero matrix of appropriate size.

(ii) The value of the knock-out option $U_\xi(t, x) = \mathbb{E}_{t,x}\{\mathrm{e}^{-r(T\wedge\widehat{\tau})}\xi(X_{T\wedge\widehat{\tau}})\}$ with maturity T, payoff function $\xi\colon \mathbb{G} \to \mathbb{R}_+$, and knock-out set $\widehat{\mathbb{G}}^c$, with

$$\widehat{\tau} = \inf\{t \in \mathbb{R}_+ : X_t \notin \widehat{\mathbb{G}}\},$$

is given by (as shown in [23])

$$U_\xi(t, x) = [\exp((T - t)\widetilde{\Lambda}_r)\xi](x),$$

where we denote by $\widetilde{\Lambda}_r$ the (sub)generator matrix

$$\widetilde{\Lambda}_r(x, y) = \begin{cases} \Lambda(x, y) - r & \text{if } x \in \widehat{\mathbb{G}},\ x = y, \\ \Lambda(x, y) & \text{if } x \in \widehat{\mathbb{G}},\ y \in \mathbb{G}, x \neq y, \\ 0 & \text{if } x \in \widehat{\mathbb{G}}^c,\ y \in \mathbb{G}. \end{cases}$$

To see that this is the case, the key observation is that the barrier option in question is a European-type option with the underlying given by the stopped process $X_{\cdot\wedge\widehat{\tau}}$, which is itself a Markov chain with generator $\widetilde{\Lambda}_0$ (the (sub)generator $\widetilde{\Lambda}_r$ is obtained when the discounting rate r is also included). Note that since $t \to \exp(tX)$ is smooth, the value-function $U_\xi(t, x)$ is smooth as a function of t.

6. Convergence

Next, we show that the convergence of a sequence of Markov chains carries over to the convergence of the corresponding American option values. We will assume that the price process S is a Markov process with state-space $\mathbb{R}_+$ that is defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$, where $\mathbf{F} = \{\mathcal{F}_t\}_{t\ge 0}$ denotes the standard filtration generated by S and Ω

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

denotes the Skorokhod space of right-continuous functions with left-hand limits that map $\mathbb{R}_+$ to $\mathbb{R}$. We take the interest rate and dividend yield to be constants equal to r and d, respectively, and assume in this section that the discounted price process $\{e^{-\gamma t}S_t\}_{t\geq 0}$ with $\gamma = r - d$ is a square-integrable martingale. In addition, we assume that S is a Feller process that solves the stochastic differential equation given by

$$\frac{dS_t}{S_{t-}} = \gamma\, dt + \sigma(S_{t-})\, dW_t + p(dt \times dx), \qquad t > 0,$$

with $S_0 = s > 0$, where W denotes a Wiener process and p denotes a compensated random measure with compensator given by the random measure $\nu(S_{t-}, dz)\, dt$, where, for every $x \in \mathbb{R}_+$, $\nu(x, dy)$ is a measure with support in $(-1, \infty)$ satisfying the integrability condition

$$\int_{(-1,\infty)} |y|^2 \nu(x, dy) < \infty.$$

The value-function $v\colon [0, T] \times \mathbb{R}_+ \to \mathbb{R}_+$ of the American option with payoff $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ on the underlying process S is denoted by

$$v(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}(\boldsymbol{F})} \mathbb{E}_{t,x}\{e^{-r(\tau - t)}\phi(S_\tau)\}, \qquad (t, x) \in [0, T] \times \mathbb{R}_+,$$

with $\mathbb{E}_{t,x}\{\,\cdot\,\} = \mathbb{E}\{\,\cdot\,|S_t = x\}$ and the set $\mathcal{T}_{t,T}(\boldsymbol{F})$ equal to the collection of $\boldsymbol{F}$-stopping times taking values between t and T. The Bermudan option, which is an American-type option for which the epoch of exercise is restricted to take values in the grid $\mathbb{T}$ given by

$$\mathbb{T} = \{i\Delta\colon i = 0, \dots, M\} \quad \text{with } \Delta = \frac{T}{M}, \tag{6.1}$$

is a closely related derivative security, with value-function $v^M\colon [0, T] \times \mathbb{R}_+ \to \mathbb{R}_+$ given by

$$v^M(t, x) = \sup_{\tau \in \mathcal{T}^M_{t,T}(\boldsymbol{F})} \mathbb{E}_{t,x}\{e^{-r(\tau - t)}\phi(S_\tau)\}, \qquad (t, x) \in [0, T] \times \mathbb{R}_+,$$

where $\mathcal{T}^M_{t,T}(\boldsymbol{F})$ denotes the collection of $\boldsymbol{F}$-stopping times taking values in the grid $\mathbb{T}$ intersected with the interval $[t, T]$.

Let $X^{(n)}$ denote a sequence of Markov chains, such that $\{e^{-\gamma\cdot}X^{(n)}_\cdot\}$ are square-integrable martingales, that is defined on the measurable space $(\Omega, \mathcal{F})$ and converges weakly to the Feller process S, where the weak convergence is in the Skorokhod J_1 topology (see, e.g. [14]). Let $V^{(n),M}$ and $V^{(n)}$ denote the value-functions of a Bermudan option with M equidistant exercise times and an American option, both with an underlying price process given by the Markov chain $X^{(n)}$. Below, we show that as n and M tend to ∞, both $V^{(n)}(0, x)$ and $V^{(n),M}(0, x)$ tend to the value $v(x)$ of the American option when the spot S_0 is equal to x. More precisely, we assume that the subsequent grids $(\mathbb{G}^{(n)})_{n\in\mathbb{N}}$ are all nested (i.e. $\mathbb{G}^{(n)}$ is contained in $\mathbb{G}^{(n+1)}$ for any positive integer n) and that the union $\cup_{n\in\mathbb{N}}\mathbb{G}^{(n)}$ is dense in $\mathbb{R}$, and consider the following convergence of a sequence of functions $f^{(n)}\colon \mathbb{G}^{(n)} \to \mathbb{R}$ to a function $f\colon \mathbb{R} \to \mathbb{R}$:

$$f^{(n)} \xrightarrow{\mathbb{G}} f \quad \Longleftrightarrow \quad \text{for all } m \in \mathbb{N},\ \text{for all } x \in \mathbb{G}^{(m)}\colon \lim_{n\to\infty, n\geq m} f^{(n)}(x) \to f(x).$$

Convergence is established under the condition that the functions

$$t \mapsto \mathbb{E}_{0,x}\{\langle e^{-\gamma\cdot}S_\cdot\rangle_t\}, \qquad t \mapsto \mathbb{E}_{0,x}\{\langle e^{-\gamma\cdot}X^{(n)}_\cdot\rangle_t\} \tag{6.2}$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

are Lipschitz-continuous on $[0, T]$ with Lipschitz constants given by $C^2(c_1x + c_2)^2$ and $D(n)^2(d_1x + d_2)^2$ for some $C, D(n), c_1, c_2, d_1, d_2 \in \mathbb{R}_+$, such that $\sup_{n\in\mathbb{N}} D(n)$ is finite, where, for any square-integrable martingale M', $\langle M'\rangle$ denotes its predictable quadratic variation. These conditions are satisfied by many of the Markov processes S used in financial modelling, and appropriately chosen approximating Markov chains $X^{(n)}$.

Theorem 6.1. *Assume that ϕ is Lipschitz-continuous and that the functions in (6.2) are Lipschitz-continuous with respective Lipschitz constants given by $C^2(c_1x+c_2)^2$ and $D(n)^2(d_1x+d_2)^2$ for some $C, D(n), c_1, c_2, d_1, d_2 \in \mathbb{R}_+$, where $\sup_{n\in\mathbb{N}} D(n)$ is finite. The following hold:*

(i) $V^{(n),M}(0,\cdot) \xrightarrow{\mathbb{G}} v^M(0,\cdot)$ *as* $n \to \infty$ *for any* $M \in \mathbb{N}$,

(ii) $V^{(n),M}(0,\cdot) \xrightarrow{\mathbb{G}} v(0,\cdot)$ *as* $\min\{n, M\} \to \infty$,

(iii) $V^{(n)}(0,\cdot) \xrightarrow{\mathbb{G}} v(0,\cdot)$ *if* $n \to \infty$.

Proof. We first prove the following claim: for any $n \in \mathbb{N}$, there exist constants $\widetilde{C}(x)$ and $\widetilde{D}(n,x)$ such that, for all $M \in \mathbb{N}$,

$$|v^M(0,x) - v(0,x)| \le \frac{\widetilde{C}(x)}{\sqrt{M}}, \qquad |V^{(n),M}(0,x) - V^{(n)}(0,x)| \le \frac{\widetilde{D}(n,x)}{\sqrt{M}}. \tag{6.3}$$

We will only prove this claim when the underlying is given by S as the proof of the case that the underlying is a Markov chain is analogous.

Observe that the collection of stopping times of the form $\tau_M = \inf\{s \ge \tau : s \in \mathbb{T}\}$ for $\tau \in \mathcal{T}_{0,T}(\boldsymbol{F})$ is equal to the set $\mathcal{T}^M_{0,T}(\boldsymbol{F})$. By an application of the triangle inequality, we find that

$$\begin{aligned}
|v(0,x) - v^M(0,x)| &\le \sup_\tau \mathbb{E}_{0,x}\{|\mathrm{e}^{-r\tau}\phi(S_\tau) - \mathrm{e}^{-r\tau_M}\phi(S_{\tau_M})|\} \\
&\le \sup_\tau \mathbb{E}_{0,x}\{|(\mathrm{e}^{-r\tau} - \mathrm{e}^{-r\tau_M})\phi(S_\tau)| + |\mathrm{e}^{-r\tau_M}(\phi(S_\tau) - \phi(S_{\tau_M}))|\} \\
&\le \frac{1}{M}c(x) + K \sup_\tau \mathbb{E}_{0,x}\{|S_\tau - S_{\tau_M}|\},
\end{aligned}$$

where the suprema are taken over the set $\mathcal{T}_{t,T}(\boldsymbol{F})$ of $\boldsymbol{F}$-stopping times taking values in the interval $[t, T]$ and we used, by the triangle inequality and the Lipschitz continuity of ϕ,

$$\sup_\tau \mathbb{E}_{0,x}\{T r \mathrm{e}^{-r\tau}|\phi(S_\tau)|\} \le Tr(\phi(x) + 2Kx) := c(x),$$

where K is the Lipschitz constant. By the strong Markov property of S and the triangle inequality the expectation on the right-hand side can be estimated by

$$\mathbb{E}_{0,x}\{|S_\tau - S_{\tau_M}|\} \le \mathbb{E}_{0,x}\{\mathbb{E}_{0,S_\tau}[|S_0 - S_{\tau_M\circ\theta_\tau}|]\}. \tag{6.4}$$

Another application of the triangle inequality yields the estimate

$$\begin{aligned}
\mathbb{E}_{0,s}\{|S_0 - S_{\tau_M\circ\theta_\tau}|\} &\le \mathbb{E}_{0,s}\{|S_0 - \mathrm{e}^{-\gamma(\tau_M\circ\theta_\tau)}S_{\tau_M\circ\theta_\tau}|\} + \mathbb{E}_{0,s}\{|\mathrm{e}^{-\gamma(\tau_M\circ\theta_\tau)} - 1|S_{\tau_M\circ\theta_\tau}\} \\
&:= e_1(s) + e_2(s)
\end{aligned} \tag{6.5}$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

for any nonnegative s. An application of Doob's optional stopping theorem to the càdlàg martingale $M' = \{M'_t = \mathrm{e}^{-\gamma t} S_t\}_{t\in[0,T]}$ implies that $e_2(s)$ can be bounded by

$$e_2(s) \le \frac{|\gamma T|\mathrm{e}^{\gamma^+ T/M}}{M}\mathbb{E}_{0,s}\{\mathrm{e}^{-\gamma(\tau_M\circ\theta_\tau)} S_{\tau_M\circ\theta_\tau}\} = \frac{|\gamma T|\mathrm{e}^{\gamma^+ T/M}}{M}s.$$

Another application of Doob's optional stopping theorem implies that the following bound holds for any $s \in \mathbb{R}_+$:

$$\mathbb{E}_{0,s}\{e_2(S_\tau)\} \le \frac{|\gamma T|\mathrm{e}^{\gamma^+ T/M}}{M}\mathrm{e}^{\gamma^+ T}\mathbb{E}_{0,s}\{\mathrm{e}^{-\gamma\tau} S_\tau\} = \frac{|\gamma T|\mathrm{e}^{\gamma^+ T/M}}{M}\mathrm{e}^{\gamma^+ T}s. \tag{6.6}$$

By an application of Doob's L^2-inequality to the martingale M' and the Lipschitz continuity, we find that

$$\begin{aligned} e_1(s) &\le \mathbb{E}_{0,s}\Big\{\sup_{s\colon s<T/M} |\mathrm{e}^{-\gamma s} S_s - S_0|\Big\} \\ &\le 4\mathbb{E}_{0,s}\{|\mathrm{e}^{-\gamma T/M} S_{T/M} - S_0|^2\}^{1/2} \\ &= 4(\mathbb{E}_{0,s}\{\langle \mathrm{e}^{-\gamma\cdot} S_\cdot\rangle_{T/M}\})^{1/2} \\ &\le 4\frac{T^{1/2}}{M^{1/2}} C(c_1 s + c_2) \end{aligned}$$

for $s \in \mathbb{R}_+$. Since M' is a martingale, we have

$$\mathbb{E}_{0,s}\{e_1(S_\tau)\} \le 4\frac{T^{1/2}}{M^{1/2}} C(c_1\mathrm{e}^{\gamma^+ T}s + c_2) \tag{6.7}$$

for any $s \in \mathbb{R}_+$. By combining (6.4)–(6.7), it follows that (6.3) holds with

$$\widetilde{C}(x) = 4KT^{1/2}C(c_1\mathrm{e}^{\gamma^+ T}x + c_2) + K|\gamma T|\mathrm{e}^{2\gamma^+ T}x + c(x).$$

Next, we turn to the proof of the three assertions.

Proof of (i). By extending the probability space if necessary, we may assume that the processes S and $(X^{(n)})_n$ are all defined on a single probability space.

Denote by $\boldsymbol{H}$ the filtration generated by the process $\{S, X^{(n)}, n \in \mathbb{N}\}$ and by $\widetilde{\mathcal{T}}^M_{t,T}$ the collection of $\boldsymbol{H}$-stopping times taking values in the set $[t, T]$ intersected with the grid $\mathbb{T}$. We may write

$$v^M(0,x) = \sup_{\tau\in\widetilde{\mathcal{T}}^M_{0,T}} \mathbb{E}_{0,x}\{\mathrm{e}^{-r\tau}\phi(S_\tau)\}, \qquad V^{(n),M}(0,x) = \sup_{\tau\in\widetilde{\mathcal{T}}^M_{0,T}} \mathbb{E}_{0,x}\{\mathrm{e}^{-r\tau}\phi(X^{(n)}_\tau)\}.$$

We have, by the triangle inequality and the Lipschitz continuity of ϕ (with Lipschitz constant K),

$$\begin{aligned} |v^M(0,x) - V^{(n),M}(0,x)| &\le \sup_{\tau\in\widetilde{\mathcal{T}}^M_{0,T}} \mathbb{E}_{0,x}\{\mathrm{e}^{-r\tau}|\phi(X^{(n)}_\tau) - \phi(S_\tau)|\} \\ &\le K \sup_{\tau\in\widetilde{\mathcal{T}}^M_{0,T}} \mathbb{E}_{0,x}\{|S_\tau - X^{(n)}_\tau|\} \qquad (6.8a) \\ &\le K\mathbb{E}_{0,x}\Big\{\sup_{t\in\mathbb{T}} |X^{(n)}_t - S_t|\Big\}, \qquad (6.8b) \end{aligned}$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

where in the last two lines we used that any stopping time τ in the set $\widetilde{\mathcal{T}}^{M}_{0,T}$ takes values in the grid $\mathbb{T}$. As, by assumption, $X^{(n)}$ converges weakly to S in the Skorokhod topology as $n \to \infty$, it follows that $X^{(n)}_t$ converges to S_t in distribution as $n \to \infty$ for any fixed $t \in \mathbb{T}$. The Skorokhod representation theorem implies that, for any given $t \in \mathbb{T}$, there exists a probability space carrying random variables $\tilde{X}^{(n)}_t$, $n \in \mathbb{N}$, and $\tilde{S}_t$ that have the same distribution as ${}^{(n)}_{t}$ and S_t, respectively, such that $\tilde{X}^{(n)}_t$ converges almost surely to $\tilde{S}_t$ as $n \to \infty$. The uniform integrability of the collection $(X^{(n)}_t, S_t, t \in \mathbb{T}, n \in \mathbb{N})$ (which is in turn a consequence of the fact that $C(x) + \sup_n D(n, x)$ is finite) thus implies that

$$\mathbb{E}_x\{|S_t - X^{(n)}_t|\} \to 0 \quad \text{as } n \to \infty \tag{6.9}$$

for any $t \in \mathbb{T}$, which also implies that the supremum in (6.8a) and (6.8b) converges to 0 since $\mathbb{T}$ contains M elements. The proof of (i) is completed by combining (6.8a) and (6.8b), and (6.9).

Proof of (ii) and (iii). The triangle inequality implies that the differences between $V^{(n)}(0, x)$, and $v(0, x)$, and $V^{(n),M}(0, x)$ and $v(0, x)$ can be estimated as

$$\begin{aligned} |V^{(n)}(0, x) - v(0, x)| &\le |V^{(n)}(0, x) - V^{(n),M}(0, x)| + |V^{(n),M}(0, x) - v(0, x)|, \\ |V^{(n),M}(0, x) - v(0, x)| &\le |V^{(n),M}(0, x) - v^{M}(0, x)| + |v^{M}(0, x) - v(0, x)|. \end{aligned} \tag{6.10}$$

Let $\varepsilon > 0$ be arbitrary. By virtue of (6.3) and the fact that $\sup_n D(n, x)$ is finite, it follows that there exists an M_ε such that, for all $M \ge M_\varepsilon$ and for all $n \in \mathbb{N}$,

$$\max\Big\{|v^{M}(0, x) - v(0, x)|, \sup_{n\in\mathbb{N}} |V^{(n)}(0, x) - V^{(n),M}(0, x)|\Big\} \le \varepsilon. \tag{6.11}$$

Fixing an M larger than M_ε, (i) implies that there exists an N_ε such that

$$|V^{(n),M}(0, x) - v^{M}(0, x)| \le \varepsilon \quad \text{for all } n \ge N_\varepsilon.$$

Combining this estimate with (6.10) and (6.11) yields the estimates $|V^{(n)}(0, x) - v(0, x)| \le 3\varepsilon$ and $|V^{(n)}(0, x) - V^{(n),M}(0, x)| \le 2\varepsilon$. Since ε was arbitrary, the statements in (ii) and (iii) follow.

7. Numerical illustrations

To provide an illustration of the effectiveness of the method we report the results of the approximation of the value of the American put option by the free-boundary approach (Algorithm 1, which we will refer to as 'FB'). The algorithm for the pricing of American options takes as input a Markov chain X that closely approximates the Feller process S which is constructed by suitably specifying its state-space and generator matrix: the state-space will be taken to be nonuniform with higher density in relevant areas (e.g. around the spot value S_0 and the strike K, in the case of a put option) and the generator matrix is chosen so as to match the first two instantaneous moments of S. The smallest and largest points of the state-space are taken sufficiently small and large, respectively, to guarantee that the truncation error is negligible at the level of accuracy that is considered in the examples below (these levels were determined after some numerical experimentation). Along these lines, an algorithm for the construction of a Markov chain was developed in [23] which we will deploy in the numerical illustrations below. By way of comparison, we also report the results of the dynamic programming algorithm

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

that proceeds by first approximating the American option by a Bermudan option by restricting the possible exercise times to a finite set and subsequently valuing the Bermudan option under the Markov chain X according to the well-known dynamic programming procedure. (This algorithm is referred to as the 'DP' algorithm and a description in the current Markov chain setting is presented in Appendix A). Additional numerical examples can be found in [12].

7.1. The CEV-Kou model

We consider the valuation of the American put option under the jump diffusion that evolves according to the stochastic differential equation

$$\frac{\mathrm{d}S_t}{S_{t-}} = \left(r - d - \lambda\xi\left(\frac{S_{t-}}{S_0}\right)^{\beta}\right)\mathrm{d}t + \left(\frac{S_{t-}}{S_0}\right)^{\beta}\mathrm{d}L_t,$$

$$L_t = \sigma W_t + \sum_{i=1}^{N_t}(\mathrm{e}^{K_i} - 1), \qquad t > 0, \quad S_0 = s > 0,$$

where W is a Brownian motion, N a Poisson process, and the K_i are independent random variables following a double exponential distribution given by

$$f_K(k) = p\lambda_p \mathrm{e}^{-\lambda_p k}\,\mathbf{1}_{(0,\infty)}(k) + (1-p)\lambda_m \mathrm{e}^{\lambda_m k}\,\mathbf{1}_{(-\infty,0)}(k), \qquad k \in \mathbb{R},$$

with $\lambda_p > 0$, $\lambda_m > 0$, and $p \in [0, 1]$. The parameter ξ is given by

$$\xi = \mathbb{E}\{\mathrm{e}^{K_1} - 1\} = \frac{p\lambda_p}{\lambda_p - 1} + \frac{(1-p)\lambda_m}{\lambda_m + 1} - 1.$$

The processes W and N and the collection of random variables $\{K_i,\ i \in \mathbb{N}\}$ are assumed to be mutually independent.

The model under consideration is a combination of the Kou model [15], a geometric Lévy process with double exponential jumps (obtained by setting $\beta = 0$), and the constant elasticity of variance (CEV) model [9], a diffusion with local volatility function given by a power (obtained by taking $\lambda = 0$). In particular, taking $\lambda = \beta = 0$, yields the geometric Brownian motion (GBM) model. This model, which we refer to as the CEV-Kou model, has an infinitesimal generator that acts on $f \in C_c^2(\mathbb{R}_+)$ as

$$\mathcal{L}f(x) = \mathcal{L}_D f(x) + \mathcal{L}_J f(x), \qquad x \in \mathbb{R}_+,$$

$$\mathcal{L}_D f(x) = (r-d)xf'(x) + \frac{\sigma^2}{2}\left(\frac{x}{S_0}\right)^{2\beta} x^2 f''(x),$$

$$\mathcal{L}_J f(x) = \int_{(-1,\infty)} [f(x(1+y)) - f(x) - f'(x)xy] f_K(\log y)\frac{\mathrm{d}y}{y}.$$

The results obtained by deploying the DP and FB algorithms are reported in Figures 3 and 4 and Table 1. Figure 3(a) shows the absolute error for the dynamic programming problem for a varying number of exercise times with fixed size of the state-space. The slope of the line in Figure 3(a) is approximately -1, which corresponds to a linear decay of the error of the DP method in $1/M$ if the number of states is fixed, where M is the number of time steps.

In Figure 3(b) we show the absolute error for the FB and DP methods with a fixed number of exercise times. We observed that the outcomes of the FB method appear to converge slightly faster than those of the DP method, but at the expense of longer execution times. Figure 3(b)

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

TABLE 1: The values of American put options under the CEV-Kou model with model parameters $r = 0.05$, $\sigma = 0.2$, $p = 0.3$, $\lambda_p = 50$, $\lambda_m = 25$, and $\lambda = 3$, obtained by using the FB and DP methods. The parameter β is given in the table, and the option parameters are $K = 100$, $S_0 = 100$, and $T = 1$.

	Size N			
	$\beta = -1$		$\beta = -3$	
	200	400	200	400
DP $M = 3200$	6.6926	6.6957	6.6576	6.6609
$M = 6400$	6.6926	6.6958	6.6577	6.6610
FB	6.6927	6.6958	6.6578	6.6611

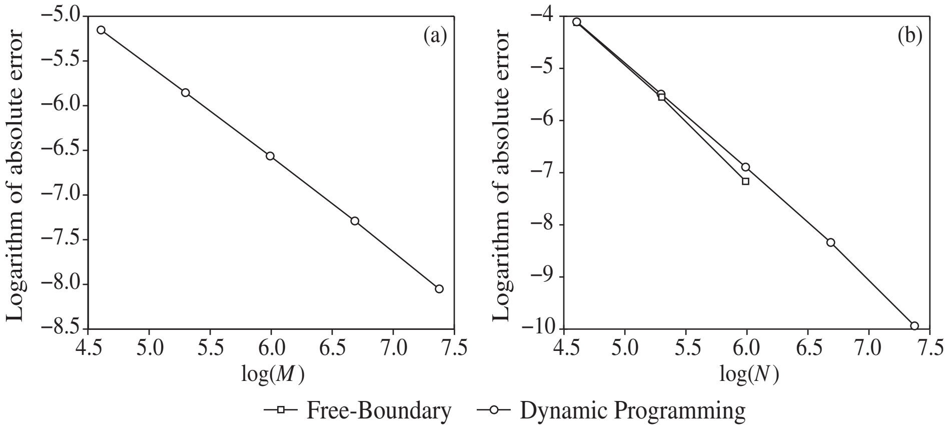

FIGURE 3: (a) The absolute error of the American put option values generated by the DP method for a varying number of exercise times M using a Markov chain with state-space of fixed size $N = 1600$. As a reference, a value is taken at the outcome of the DP method for $M = 12800$ exercise times. The Markov chain is an approximation to the CEV-Kou model with model parameters given by $r = 0.05$, $d = 0$, $\sigma = 0.2$, $\beta = -1$, $p = 0.3$, $\lambda_p = 50$, $\lambda_m = 25$, and $\lambda = 3$. The option parameters are fixed to be equal to $K = 100$ (strike), $S_0 = 100$ (spot), and $T = 1$ (maturity). (b) The absolute error of the American put option prices with the same parameters under the same model as in (a) for varying sizes N of the state-space of the Markov chain for the FB method and DP method with $M = 6400$. In both (a) and (b) the reference values for the computation of the errors of the values generated by the FB and DP methods are taken as equal to the outcomes generated by these two methods with $N = 800$ and $N = 3200$ states, respectively.

appears to show a quadratic speed of convergence in $1/N$ with N the cardinality of the state-space $\mathbb{G}$. Figure 4 contains the execution times for the outcomes obtained by the FB and DP algorithms for a varying number of states N of the approximating Markov chain, showing that the DP algorithm is the faster of the two. We observed that the change in execution time when varying the number of exercise times is very small. One explanation for this small change is that the bulk of the computational effort is in calculating the matrix exponential $\exp(\Delta\Lambda)$, and it appears that the time to calculate $\exp(\Delta\Lambda)$ is only marginally affected by the size of Δ, and decreasing Δ often results in slightly faster calculations. For $\beta = 0$, the CEV-Kou model reduces to the Kou model. In Table 2 we compare the results obtained using the DP and FB

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

TABLE 2: Displayed are American put option prices under the Kou model (which is equal to the CEV-Kou model with $\beta = 0$). The final two columns are obtained from [16]. In all cases it is assumed that the spot is $S_0 = 100$, the maturity is $T = 1$, the interest rate is $r = 0.06$, the volatility is $\sigma = 0.2$, and the probability of an upward jump is $p = 0.6$, with the remaining parameters as given in the table. We employ a Markov chain with state-space of size $N = 400$, and for the DP algorithm we used $M = 3200$ exercise times.

K	λ	λ_p	λ_m	FB	DP	Kou binomial	Kou approximation
90	3	50	25	2.6709	2.6707	2.66	2.72
90	3	50	50	2.4568	2.4566	2.46	2.51
90	7	25	50	3.2282	3.2280	3.24	3.29
90	7	50	50	2.6662	2.6660	2.66	2.72
100	3	50	25	6.2700	6.2698	6.26	6.29
100	3	50	50	6.0120	6.0118	6.01	6.03
100	7	25	50	7.0524	7.0522	7.07	7.09
100	7	50	50	6.2891	6.2889	6.28	6.31
110	3	50	25	12.0559	12.0557	12.04	12.00
110	3	50	50	11.8442	11.8440	11.84	11.78
110	7	25	50	12.8296	12.8294	12.85	12.79
110	7	50	50	12.0928	12.0926	12.08	12.03

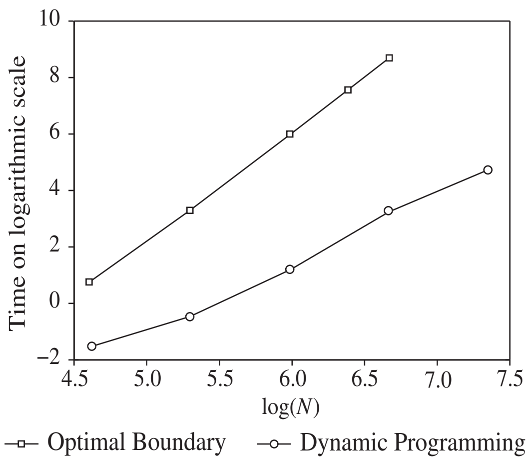

FIGURE 4: Displayed are the execution times for the computation of the American put option deploying the FB and the DP ($M = 6400$ exercise times) methods for various sizes N of the approximating Markov chain. The option parameters are fixed and taken to be $K = 100$ (strike), $S_0 = 100$ (spot), and $T = 1$ (maturity). The underlying price process follows a CEV-Kou model with parameters $r = 0.05$, $d = 0$, $\sigma = 0.2$, $\beta = -1$, $p = 0.3$, $\lambda_p = 50$, $\lambda_m = 25$, and $\lambda = 3$. Computations were carried out in Matlab® on a laptop with Intel® Core™ Duo T2500 2GHz.

methods in the case $\beta = 0$ with those reported in [16]. Note that although the results are reported in [16] for an interest rate equal to $r = 0.05$, we match the numbers in [16] by using the value $r = 0.06$. We believe that this is a misprint in [16]. For $\lambda = 0$, the CEV-Kou model reduces to the CEV model and in Table 3 we report the outcomes of the FB and DP methods in the $\beta = 0$ and $\beta = -\frac{1}{3}$ cases, and the results obtained in [27] using a finite difference method.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

TABLE 3: Value of the at-the-money American put option with strike $S_0 = K = 100$. In the upper part of the table the underlying is a GBM ($\beta = 0$) with parameter values taken from [7] ($r = 0.1$, $\delta = 0$, $\sigma = 0.3$, and maturity $T = 1$). The row CR refers to Carr's [7] randomisation algorithm with the number of randomisation steps taken equal to 15 (using Richardson's extrapolation). In the lower part of the table the underlying is given by the CEV model with parameters taken from [27] ($r = 0.05$, $q = 0$, $\sigma = 0.2$, $\beta = -\frac{1}{3}$, and maturity $T = 0.5$). The row 'WZ' refers to results obtained by [27] using a finite difference scheme. The row 'Binomial' refers to the outcomes of a binomial tree algorithm with 2000 time steps (above) and 5000 time steps (below). For the Markov chain for the DP method and FB method, 'Size' denotes the size of the state-space of the Markov chain.

	$\beta = 0$		
	$N = 200$	$N = 400$	$N = 800$
DP $M = 3200$	8.3316	8.3359	8.3370
$M = 6400$	8.3318	8.3361	8.3371
FB	8.3320	8.3363	8.3373
CR			8.3371
Binomial			8.3378
	$\beta = -\frac{1}{3}$		
	$N = 400$	$N = 600$	$N = 800$
DP $M = 1600$	4.6488	4.6490	4.6491
$M = 3200$	4.6488	4.6491	4.6491
FB	4.6489	4.6491	4.6492
WZ			4.6489
Binomial			4.6491

Appendix A. Dynamic programming algorithm

A Bermudan option with payoff function ϕ and a finite set of admissible exercise times $\mathbb{T} \subset [0, T]$ is a derivative security that may be exercised at any time $\tau \in \mathbb{T}$ yielding payoff $\phi(X_\tau)$. For the ease of presentation, we restrict ourselves to the case of an equidistant grid given in (6.1) with mesh size $\Delta = T/M$. The value $V(t, x)$ of the Bermudan option at time $t \in \mathbb{T}$ in case we have $\{X_t = x\}$ is given by

$$V(t, x) = \max_{\tau \in \mathcal{T}_{t,T}(\Delta, \mathbb{G})} \mathbb{E}_{t,x}\{\mathrm{e}^{-r(\tau - t)}\phi(X_\tau)\}$$

for $t \in \mathbb{T}$, and $x \in \mathbb{G}$, where $\mathcal{T}_{t,T}(\Delta, \mathbb{G})$ is the set of $\boldsymbol{G}$-stopping times τ taking values in $[t, T] \cap \mathbb{T}$, where $\boldsymbol{G} = \{\mathcal{G}_t, t \in [0, T]\}$ denotes the filtration generated by the Markov chain X. At any time $t \in \mathbb{T}$, the holder of the Bermudan option has the choice between immediately exercising or continuing to wait. The former results in a payoff of $\phi(X_t)$, while in the latter case the expected reward of postponing exercise, assuming that the holder continues to follow an optimal strategy from time t to maturity, is $\mathbb{E}_{t,X_t}\{\mathrm{e}^{-r\Delta}V(t + \Delta, X_\Delta)\}$. Thus, for any $t \in \mathbb{T}$, the value $V(t, x)$ is at least equal to the larger of $\phi(x)$ and $\mathbb{E}_{t,x}\{\mathrm{e}^{-r\Delta}V_{t+\Delta}(X_\Delta)\}$. The DP principle states that in fact equality holds: with $V_i(x) = V(i\Delta, x)$, we have

$$V_i(x) = \max(\phi(x), \mathbb{E}_{i\Delta,x}\{\mathrm{e}^{-r\Delta}V_{i+1}(X_{(i+1)\Delta})\})$$

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

for $i = 0, \dots, M-1$, and $x \in \mathbb{G}$. Noting that in view of the form of the semigroup in (2.1), we have

$$\mathbb{E}_{t,x}\{\mathrm{e}^{-r\Delta} V_{i+1}(X_\Delta)\} = [\exp(\Delta\Lambda^{(r)})V_{i+1}](x).$$

By deploying the DP principle, we obtain the following recursive procedure to compute the values of $V_i(x)$ ranging over all initial values $x \in \mathbb{G}$ and all time-steps $i = 0, \dots, M$.

Algorithm 2. (*Procedure to compute the value of a Bermudan option.*)

```
set △ ← T/M
set V ← O ∈ ℝ^{N×(M+1)}
set V( : , M + 1) ← φ( : )
evaluate A = exp(△Λ^{(r)})
for i = M to 1
    V( : , i) ← A[V( : , i + 1)];
    V( : , i) ← max(φ( : ), V( : , i));
    i ← i − 1;
end
return V
```

Remark A.1. (i) The algorithm returns the matrix $(V_i(x), (i\Delta, x) \in \mathbb{T} \times \mathbb{G})$ of values of the Bermudan option on the time-space grid $\mathbb{T} \times \mathbb{G}$, where $V(:, i)$ denotes the ith column of the matrix V and contains the values $V_{i+1}(x)$ for $x \in \mathbb{G}$.

(ii) Note that when, as assumed above, the time-grid $\mathbb{T}$ is equidistant, the exponentiation of the matrix $\Delta\Lambda$ needs to be computed only once. If the time-grid $\mathbb{T}$ is chosen nonequidistant, the above algorithm will computationally be a good deal more expensive, since a costly exponentiation would need to be carried out at every iteration of the recursive procedure.

Appendix B. Proof of Dynkin's lemma (Lemma 2.1)

Proof. First, assume that, in addition to the stated assumptions, the map $F(\cdot, x)\colon [0, T] \to \mathbb{R}$ is continuously differentiable for every $x \in \mathbb{G}$. An application of Itô's lemma to the semimartingale $\{\mathrm{e}^{-rt}F(t, X_t)\}_{t\in[0,T]}$ shows that the process $\{M_t\}_{t\in[0,T]}$ with

$$M_t = \mathrm{e}^{-rt}F(t, X_t) - F(0, X_0) - \int_0^t \mathrm{e}^{-rs}\left[\frac{\partial F}{\partial t} + (\Lambda F) - rF\right](s, X_s)\,\mathrm{d}s$$

is a local martingale. In view of the assumptions on F and Λ it follows that M is in fact a uniformly integrable martingale. An application of Doob's optional stopping theorem implies that for every $\boldsymbol{G}$-stopping time τ taking values in $[t, T]$, we have $\mathbb{E}_{t,x}\{M_\tau\} = 0$, so (2.3) holds.

Next, assume that F is as stated in the lemma, with density f. Since the set $\mathcal{G}$ of functions $G\colon [0, T] \times \mathbb{G} \to \mathbb{R}$ that is continuously differentiable at $t \in [0, T]$ for every $x \in \mathbb{G}$ is dense in the set of continuous real-valued functions with domain $[0, T] \times \mathbb{G}$, there exists a sequence of functions $(G_n)_n$ in $\mathcal{G}$ that almost everywhere converges to F. An application of Lebesgue's dominated convergence theorem which is justified by the facts that F is bounded and Λ has uniformly bounded diagonal (cf. Remark 2.1) shows that (2.3) is true under the stated assumptions.

Acknowledgements

This research was supported by EPSRC grant no. EP/D039053. We thank Aleksandar Mijatović for useful conversations.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press

References

[1] Abergel, F. and Jedidi, A. (2011). A mathematical approach to order book modelling. In *Econophysics of Order-Driven Markets*, Springer, Milan, pp. 93–107.
[2] Ahn, J. and Song, M. (2007). Convergence of the trinomial tree method for pricing European/American options. *Appl. Math. Comput.* **189,** 575–582.
[3] Alili, L. and Kyprianou, A. E. (2005). Some remarks on first passage of Lévy processes, the American put and pasting principles. *Ann. Appl. Prob.* **15,** 2062–2080.
[4] Boyarchenko, S. I. and Levendorskiĭ, S. Z. (2000). Option pricing for truncated Lévy processes. *Internat. J. Theoret. Appl. Finance* **3,** 549–552.
[5] Boyarchenko, S. and Levendorskiĭ, S. (2002). Perpetual American options under Lévy processes. *SIAM J. Control Optimization* **40,** 1663–1696.
[6] Boyarchenko, S. I. and Levendorskiĭ, S. Z. (2007). *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*. Springer, Berlin.
[7] Carr, P. (1998). Randomization and the American put. *Rev. Financial. Stud.* **11,** 597–626.
[8] Chung, K. L. and Walsh, J. B. (2005). *Markov Processes, Brownian Motion, and Time Symmetry*, 2nd edn. Springer, New York.
[9] Cox, J. (1996). Notes on option pricing I: constant elasticity of variance diffusions. *J. Portfolio Manag.* **22,** 15–17.
[10] Cox, J. C., Ross, S. A. and Rubinstein, M. (1979). Option pricing: a simplified approach. *J. Financial Econom.* **7,** 229–263.
[11] Detemple, J. (2006). *American-Style Derivatives*. Chapman & Hall/CRC, Boca Raton, FL.
[12] Eriksson, B. (2013). On the valuation of barrier and American options in local volatility models with jumps. Doctoral Thesis, Imperial College London.
[13] Jacka, S. D. (1991). Optimal stopping and the American put. *Math. Finance* **1,** 1–14.
[14] Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.
[15] Kou, S. G. (2002). A jump-diffusion model for option pricing. *Manag. Sci.* **48,** 1086–1101.
[16] Kou, S. G. and Wang, H. (2004). Option pricing under a double exponential jump diffusion model. *Manag. Sci.* **50,** 1178–1192.
[17] Kushner, H. J. (1997). Numerical methods for stochastic control problems in finance. In *Mathematics of Derivative Securities* (Publ. Newton Inst. **15**), Cambridge University Press, pp. 504–527.
[18] Kushner, H. J. and Dupuis, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd edn. Springer, New York.
[19] Lamberton, D. (1998). Error estimates for the binomial approximation of American put options. *Ann. Appl. Prob.* **8,** 206–233.
[20] Lamberton, D. and Mikou, M.A. (2012). The smooth-fit property in an exponential Lévy model. *J. Appl. Prob.* **49,** 137–149.
[21] Lamberton, D. and Mikou, M. A. (2013). Exercise boundary of the American put near maturity in an exponential Lévy model. *Finance Stoch.* **17,** 355–394.
[22] McKean, H. P., Jr. (1965). Appendix: a free boundary problem for the heat equation arising from a problem in mathematical economics. *Indust. Manag. Rev.* **6,** 32–39.
[23] Mijatović, A. and Pistorius, M. (2013). Continuously monitored barrier options under Markov processes. *Math. Finance* **23,** 1–38.
[24] Peskir, G. (2005). On the American option problem. *Math. Finance* **15,** 169–181.
[25] Peskir, G. and Shiryaev, A. (2006). *Optimal Stopping and Free-Boundary Problems*. Birkhäuser, Basel.
[26] Szimayer, A. and Maller, R. A. (2007). Finite approximation schemes for Lévy processes, and their application to optimal stopping problems. *Stoch. Process. Appl.* **117,** 1422–1447.
[27] Wong, H. Y. and Zhao, J. (2008). An artificial boundary method for American option pricing under the CEV model. *SIAM J. Numerical Anal.* **46,** 2183–2209.

https://doi.org/10.1239/aap/1435236980 Published online by Cambridge University Press